Implementing PREMIS: A Case Study of the Florida Digital Archive

Journal:	*Library Hi Tech*
Manuscript ID:	Draft
Manuscript Type:	Original Article
Keywords:	Preservation Metadata, PREMIS, Digital Preservation

I. Introduction

With increasing sophistication, scholars and practitioners are treating digital preservation as a "research subject and social phenomenon" (Seadle, 2009, p. 319), due in part to the rapid growth of digital content worthy of long-term retention and in part to the genuine technical and administrative complexity that surround the development of preservation repositories (Hedstrom, 2003). Libraries, archives, museums, and other cultural heritage institutions that accept responsibility for digital preservation also accept the challenge of implanting preservation systems that conform to emerging standards and best practices (Conway, 1996). Among the many clusters of standards that are evolving in the preservation repository space, preservation metadata poses particular implementation challenges because of the cost and complexity of implementing metadata systems that are largely unknown and unseen by users and yet form the foundation of active digital preservation management (Lavoie and Gartner, 2005). In a repository, the preservation metadata scheme is a subset of administrative and technical metadata that provides a structured way to "describe and record information needed to manage the preservation of digital resources" (PADI, 2007). Just as honey is inextricably tied to the honeycomb, preservation metadata is wed to and necessary for digital preservation.

In spite of the existence of a broad consensus in the digital preservation community on the importance of preservation metadata, few documented instances of preservation metadata systems exist and great variation prevails among the implementers in the definition of preservation metadata schemes. PREMIS, the pronounceable acronym for the PREservation Metadata: Implementation Strategies standards initiative, is one widely recognized approach to defining the elements of preservation metadata and supporting implementation in digital repositories (PREMIS Editorial Committee, 2008). Because an international digital preservation community developed the PREMIS Data Dictionary and related implementation guidelines, over a three year-period (2004-06) explicitly for use within that community, one would expect that PREMIS would have been quickly and widely adopted. This has not yet happened (Alemneh, 2008), and the reasons for slow adoption of PREMIS are not well understood.

The purpose of this case study is to describe and interpret the PREMIS implementation process in one digital repository, and in doing so provide insight into the barriers to the adoption of PREMIS. The study is based on the implementation experience at the Florida Digital Archive (FDA) of the Florida Center for Library Automation (FCLA), which is one of the most established and well-recognized preservation repositories in the United States. Through the lens of Diffusion of Innovations (DOI) and Management Science and Information Systems (MIS) models of technology implementation, the research focuses on two questions: (1) What does the adaptation stage look like when implementing PREMIS? And (2) How do developers describe the process of implementing PREMIS? A deeper understanding of the details of the PREMIS implementation process advances the goals of the digital preservation community to implement PREMIS widely in organizations that preserve digital content.

II. Literature Review

The Origins of PREMIS

Garrett and Waters (1996) first articulated the need for trustworthy digital repositories to preserve culturally significant digital objects and provide the mechanisms that minimize the threat posed by the ease by which such digital content can be altered, manipulated, corrupted, or lost. Well over a decade ago, Garrett and Waters wrote that "the integrity of information objects in the digital environment is today so fragile that as stakeholders disseminate, use, reuse, recreate and re-disseminate various kinds of digital information, they can easily, even inadvertently, destroy valuable information" (p. 19).

Garrett and Waters provide a significant contribution to the advancement of understanding digital preservation by outlining the elements of a theory of information integrity that requires the preservation of specific features of digital information, including content, fixity, reference, provenance, and context. The Consultative Committee for Space Data Systems (CCSDS) proposed the Open Archival Information System (OAIS) Reference Model and adopted Garrett and Waters' model of integrity as the centerpiece of its definition of Preservation Description Information (CCSDS, 2002, pp. 4-27-28). The CCSDS's model is not prescriptive, nor does it specify an implementation framework for preservation information in a digital archive (CCSDS, 2002, pp. 1-3).

Building explicitly on the OAIS reference model, the Online Computer Library Center/Research Libraries Group (OCLC/RLG) Working Group on Preservation Metadata expanded its conceptual framework and mapped a set of metadata elements to reflect the concepts and requirements set forth in the OAIS model (OCLC/RLG, 2002). The working group defined the role of preservation metadata within the context of "Content Information" which "consists of the Content Data Object (Physical Object or Digital Object, i.e., bits) and its associated Representation Information needed to make the Content Data Object understandable to the Designated Community" (CCSDS 2002. pp. 2-5; OCLC/RLG, 2002, p. 9). The Preservation Metadata: Implementation Strategies (PREMIS) working group took the next conceptual step in defining preservation metadata by proposing a Data Dictionary and an XML schema to support implementation of the Data Dictionary in digital archiving systems (PREMIS Working Group, 2005, p. viii).

The digital preservation community experimented with the PREMIS Data Dictionary and its implementation, identified errors, and provided feedback on ways that the Data Dictionary could be improved (PREMIS Editorial Committee, 2008). In March 2008 the PREMIS Editorial Committee released version 2.0 of the PREMIS Data Dictionary for Preservation Metadata (PREMIS Editorial Committee, 2008).

PREMIS codifies and defines preservation metadata in the form of a Data Dictionary and a schema for representing PREMIS data in XML (Caplan, 2009, p. 4). The PREMIS Data Dictionary and accompanying schema consist of *entities* and *semantic units*. PREMIS *entities* are '"things" (intellectual entities, objects, events, agents, and rights) that are considered important to talk about in the context of a digital preservation repository system" (Caplan, 2009, p. 7). *Semantic units* are "pieces of information or

knowledge related to PREMIS entities that digital repository systems need to know and should be able to export to other systems" (Caplan, 2009, p. 7).

PREMIS by itself is simply an information model that has no direct link to preservation practice unless the model is embedded in a working database system. Institutions can use the PREMIS Data Dictionary as a guide for making sure they are aware of information that members of the international digital preservation community have deemed as necessary to record for the sake of preservation. On the other hand, institutions wishing to take full advantage of the power of preservation metadata to manage their digital objects must have a digital repository management system in place that effectively utilizes the PREMIS Data Dictionary and its accompanying XML schema.

It is clear that significant progress has been made over the past thirteen years to define preservation metadata requirements and specify the structure of preservation metadata within a preservation repository. Little evidence exists, however, that PREMIS has been implemented in any significant way by preservation repository managers. As of January 2010, the PREMIS Implementation Registry lists twelve institutions (PREMIS Maintenance Activity, 2010). Of this dozen, fewer than three have fully realized preservation metadata capabilities that are compliant with the PREMIS model.

PREMIS Implementation in the Literature

The research literature on digital preservation sheds little light on the complexities and barriers of metadata system implementation. The literature on PREMIS itself is not very extensive, encompassing definitional matters (Lavoie, 2004; Caplan and Guenther, 2005; McCallum, 2005; Caplan, 2009), general recommendations on implementing PREMIS (Lee et al., 2006), intended uses of PREMIS (Knight and Hedges, 2007), descriptions of PREMIS use by specific institutions (Dappert and Enders, 2008; Pearce et al., 2008), and guidelines for using PREMIS with the Metadata Encoding and Transfer Standard (METS) (Guenther and Xie, 2007; Guenther, 2008). Only two research articles present specific data on PREMIS implementation and speculate on the barriers that may hinder implementation of PREMIS. Woodyard-Robinson (2007) focuses on how institutions implement the PREMIS Data Dictionary. Alemneh (2008) explores the barriers to adoption of PREMIS in cultural heritage institutions.

At the time Woodyard-Robinson wrote, most of the institutions she described were still in the planning stages of preservation metadata implementation; a few had started development, some had mapped their existing metadata to the PREMIS Data Dictionary; and many were planning further development to their current systems – specifically by integrating the PREMIS Data Dictionary into their systems (Woodyard-Robinson, 2007). She examined sixteen institutions and uncovered specific information about which PREMIS entities and semantic units that implementing institutions were using or were planning to use. Many institutions used *Object*, *Event*, and *Agent* PREMIS entities, while the *Rights* entity was not widely implemented. Woodyard-Robinson also finds that institutions were planning to use several semantic units.

Results from Woodyard-Robinson's study suggest that the most important decisions regarding PREMIS implementation concern the choice of entities and semantic units as well as how to gather the appropriate information that should correspond to PREMIS entities and semantic units and represent this information in a digital

preservation repository system. Woodyard-Robinson's study has little to say about the actual nature or pattern of this decision-making process.

Alemneh (2008) suggests that part of the reason why PREMIS has not yet been fully adopted by many institutions is because they were waiting for release of the second version of the Data Dictionary. The most frequently identified barriers to the adoption of PREMIS include lack of training/expertise and perceived lack of knowledge necessary to be confident in the ability to implement PREMIS (Alemneh, 2008, p. 117). The same study also finds that while many institutions have made the decision to adopt PREMIS, few have fully adopted PREMIS (Alemneh, 2008, p. 115). These findings suggest that the problem of low adoption of PREMIS may not be attributed to any barriers concerning the decision to adopt PREMIS. Instead, barriers present themselves after the decision to adopt during the implementation process itself. A more thorough understanding of the PREMIS implementation process could perhaps help mitigate or remove completely barriers to the adoption of PREMIS.

Diffusion of Innovations/Management Science & Information Systems Literature

The process of PREMIS implementation has not been described in the context of a formal model of technology adoption. . However, several process models and stage research models exist that highlight the implementation process of innovations and information technologies more generally. In Rogers' (2003) classic Five-Stage process model for organizational adoption and implementation, stages 3 through 5 (redefining, clarifying, routinizing) are helpful for understanding what goes on during an implementation process, but are too general in their description to be applicable to PREMIS implementation directly. Cooper & Zmud (1990) adapt Rogers' model to the implementation of Information Technology (IT). Their Information Technology Implementation Process model includes six stages: initiation; adoption; adaptation; acceptance; routinization; and infusion. Each of the stages of the implementation process defined in Rogers' and Cooper & Zmud's models are relevant to PREMIS implementation; but of particular importance to this study is the "adaptation" stage. Implementation of any given technology or innovation almost never fits perfectly into an organization (Berman, 1980), so full implementation is highly unlikely to occur without traversing the critically important adaptation stage.

Rogers' and Cooper & Zmud's definitions lack specificity with regard to the adaptation stage of the implementation process. Leonard-Barton (1988) more specifically defines the adaptation process as a technology adaptation cycle incorporating the following eight steps: (1) idea generation, (2) problem solving, (3) concept definition, (4) test of feasibility, (5) laboratory prototype, (6) pilot production prototype, (7) production prototype, and (8) production-ready technology. More detailed than Rogers' and Cooper & Zmud's discussions of the adaptation stage of the implementation process, Leonard-Barton's steps provide context for how adaptation might actually be executed.

Because adaptation is essential for implementation, this study seeks to address the following research questions:

1. What does the adaptation stage look like when implementing PREMIS?
2. How do developers describe the process of implementing PREMIS?

III. Methodology

The research conducted on PREMIS implementation utilized the qualitative case study analysis method (Yin, 2003). The approach involves the investigation of a specific case, "with the intent of examining an issue with the case illustrating the complexity of an issue" (Creswell 2007, p. 93). Qualitative case study analysis also involves a bounded system, the use of multiple sources of information in data collection, and the creation of a detailed and contextualized picture of a particular issue (Creswell, 2007).

As a case for study, the Florida Digital Archive is an exemplar of an organization in the throes of deciding just how to implement the PREMIS metadata model in a working system. The mission of the Florida Digital Archive is to provide a cost-effective, long-term preservation repository for digital materials in support of teaching and learning, scholarship, and research in the state of Florida. During the time this study was conducted, the FDA was fully engaged in the implementation process and attuned to the complexities of adapting the PREMIS metadata scheme. Previous studies have included FDA in the set of organizations implementing PREMIS, providing important background information for the case study (FCLA, 2006; Woodyard-Robinson, 2007; Fischer et al., 2008). The Florida Digital Archive is rich in technical expertise and human resources dedicated to PREMIS implementation, including four full-time programmers, a project manager, and a director who is in charge of implementing PREMIS. The FDA has unique expertise in PREMIS – given that the FDA's Director served on the PREMIS Working Group and is a current member of the PREMIS Editorial Committee. The use of FDA as a case informally controls for some of the major known barriers to implementation, including the lack of technical expertise, resources, and knowledge of PREMIS (Alemneh, 2008), freeing the investigation to focus on decision-making strategies.

This case study is bounded by time (two full weeks in residence during early June 2009) and place (at the Florida Digital Archive in Gainesville, Florida). Sources of information used in data collection include in-person semi-structured interviews, field observations, and prototype examples of specific implementation outcomes. The major advantage of field observations, carried out in tandem with interviews, is the opportunity to compare developers' answers with observed behavior. In addition, conducting field observations in the midst of FDA decision-making allows for the assessment of the combination of observed behavior and answers to interview questions against Diffusion of Innovations (DOI) and Management and Information Systems (MIS) literature and models – in an effort to better understand the PREMIS implementation process in the context of the FDA.

Interviews with the Director of the Florida Digital Archive established the parameters of the technical expertise marshaled for PREMIS implementation. Three of four systems developers assigned to PREMIS implementation were interviewed for the study. Even though four developers shared implementation responsibilities and exchanged information through implementation working group meetings, Developer A was generally recognized as serving in the lead because of his deeper knowledge regarding PREMIS implementation.

Utilizing a full list of PREMIS entities and semantic units (see PREMIS Data Dictionary, Version 2.0), the first author conducted semi-structured interviews in which each of the three available developers was asked to indicate which semantic units the

FDA used or did not use during implementation. Each developer was asked specifically about the nature of the PREMIS implementation process, with the question: "What is the process or steps involved when trying to implement PREMIS?" Developers were also asked if there were any challenges to the implementation of PREMIS, and if so, to explain what those challenges were and how they were addressed. Each developer was interviewed informally and separately in his or her office during normal business hours. Developers also provided examples of use of PREMIS. Consequently, an additional interview was conducted in which two developers described a prototype example they had worked on together. Each interview was tape recorded and later transcribed.

Field observations focused on the extent to which stages defined in DOI/MIS literature seem to have appeared at the FDA. Observations also focused on what happened during these stages or what evidence could be used to suggest that developers at the FDA were actively participating in certain adaptation stages. Because the PREMIS implementation process involves making decisions about which PREMIS semantic units and entities to use for a given database application, close attention was paid to any mention of developers' decisions regarding use or intended use of PREMIS entities and semantic units during FDA working group meetings. Special note was made when PREMIS entities and semantic units were mentioned in any observed conversations, based on the assumption that mention of PREMIS entities and semantic units by developers while they were working would serve as a cue for discussion of the PREMIS implementation process. Because developers were observed in their natural working environment, the research yielded unforced data on the adaptation stage of the PREMIS implementation process in much greater detail than previously covered in the literature.

IV. Findings

Findings suggest that during PREMIS implementation at the FDA developers consider what information they already collect and how closely this information relates to PREMIS concepts, entities, and/or semantic units. Based on this comparison, developers make decisions about how best to represent the information they have available in PREMIS. The data from this study provide more context for the PREMIS implementation process, i.e. what specific information the FDA compared with what specific PREMIS information to come to decisions about how to use PREMIS and what changes needed to be made in order to represent the FDA's information using PREMIS.

Interview data reveal that developers have clearly defined steps governing the PREMIS implementation process. When asked what the process for implementing PREMIS was like, specifically if this process could be typified by steps, one developer responded:

"[w]e start with our problem, from the problem, we get the information we know we need to represent, and then once we have that information, we choose the appropriate semantic units of PREMIS."

--Developer A

When asked the same question, a second developer replied:

"I have this information that I want to get across in my event, and the PREMIS Data Dictionary told me I have these tools. It was just it was just really a judgment call."

--Developer B

In both examples, the developers describe a process of clearly defined steps in which developers consider information about the FDA (its systems, practices, and data) and gather information about PREMIS (the Data Model and Data Dictionary), compare all of this information, and then make a judgment call – the outcome of which involves selection of appropriate PREMIS semantic units and entities, based on the FDA's information need for proper representation of information needed for access, management, and long term preservation of its holdings.

Interview data and prototype examples acquired in the field at the FDA provide insight into how the FDA has adapted information it requires all its affiliates to provide to PREMIS. The FDA requires all affiliates (e.g. state universities and colleges in Florida that have agreements to submit their digital objects to the FDA for safekeeping) to submit digital objects as Submission Information Packages (SIP) that are bundled files containing digital objects along with an XML document describing the digital objects (FCLA, 2006). The structure of the XML documents included in a given SIP, called SIP descriptors, must comply with the structure and format of the Metadata Encoding and Transmission Standard (METS). The METS schema is a standard for encoding descriptive, administrative, and structural metadata regarding objects within a digital repository. Since some metadata elements included in the SIP are essential for preservation functionality within the FDA, developers, as part of the PREMIS implementation process, must make decisions about how to represent in PREMIS the SIP-originated metadata provided by affiliates.

During an interview with one of the FDA's lead developers, the developer shared an example of a SIP descriptor describing a single tiff image and explained how the metadata provided in METS would be represented in PREMIS at the FDA. The following examples are parts taken from the example SIP descriptor the developer shared, followed by the developer's explanation of the parts and the developer's explanation of how the parts are being adapted in PREMIS at the FDA:

Example 1 of METS in SIP descriptor:

```
-<dmdSec ID="DMD-0">
 -<mdWrap MDTYPE="MODS">
 -<xmlData>
  -<mods>
   -<titleInfo>
    <title>The(fd)A Team</title>
   -</titleInfo>
  -<mods>
  -</mdWrap>
 -<dmdSec>
```

Explanation of Example 1 of METS in SIP descriptor:

"Starting with the first element in the METS descriptor, there is a descriptive metadata section with a MODS title and this descriptive metadata section applies to the package as a whole."

--Developer A During Interview with Developer B

Explanation of how the FDA would represent in PREMIS the information in Example 1 of METS in SIP descriptor:

"Inside of the [PREMIS semantic unit called] objectCharacteristicsExtension of the PREMIS object we can insert the MODS data. So, now we have a PREMIS object of type representation with some MODS data we took from the descriptive metadata section."

--Developer A During Interview with Developer B

Example 2 of METS in SIP descriptor:

```
-<amdSec>
 -<digiprovMD ID="DPMD-0">
 -<mdWrap MDTYPE="OTHER" OTHERMDTYPE="DAITSS">
  -<xmlData>
   -<daitss>
   -<AGREEMENT_INFO ACCOUNT="ACT" PROJECT="PRJ"/>
   -</daitss>
  -</xmlData>
  -</mdWrap>
 -</digiprovMD>
 -</amdSec>
```

Explanation of Example 2 of METS in SIP descriptor:

"The next element is an AMD section which contains some digital provenance, applying to the entire package. The type of metadata is DAITSS from the [locally created] DAITSS schema. This metadata identifies the account and project which are local. This information is semantic elements that determine which account the package belongs to and within which project underneath that account it belongs to."

--Developer A During Interview with Developer B

Explanation of how the FDA would represent in PREMIS the information in Example 2 of METS in SIP descriptor:

"We add a [PREMIS semantic unit called] objectCharacteristicsExtension to the previously created object of type representation, that applies to the whole package, and we put in that DAITSS schema in there."

--Developer A During Interview with Developer B

"Say there's a technical metadata section in the amdSec in the METS and it is referenced by a file via ADMID. We take the data inside the mdWrap because all FCLA packages up to this point use mdWrap and we put that inside of the objectCharacteristicExtension in the PREMIS object for that file."

--Developer A During Interview with Developer B

Example 3 of METS in SIP descriptor:

```
-<fileSec>
 -<fileGrp>
        -<file ID="FILE-0"
        CHECKSUM="805ae75bc4595521e350564c90a56d28"
        CHECKSUMTYPE="SHA-1">
        <FLOCAT xlink:href="ateam.tiff" LOCTYPE="URL"/>
  -</file>
 -</fileGrp>
-</fileSec>
```

Explanation of Example 3 of METS in SIP descriptor:

"Moving on, we get to the fileSec, and in the METS fileSec there is one file, ateam.tiff, and what this demonstrates is that we have the file of ID file-0, the checksum of checksum type sha-1, and then there's a location for that file, and that file location is ateam.tiff and its location type is url."

--Developer A During Interview with Developer B

Explanation of how the FDA would represent in PREMIS the information in Example 3 of METS in SIP descriptor:

"So, to represent that in PREMIS, the path we took was to make an object, a PREMIS object, of type file. Inside of objectCharacteristics, there's a section for fixity. We made a fixity element where the algorithm is sha-1 and the value is the value taken from the file checksum".

--Developer A During Interview with Developer B

Examples 1 through 3 illustrate how all the descriptive metadata affiliates submit in SIPs get adapted for PREMIS. Example 4 illustrates how structural metadata affiliates submit in SIPs get represented in PREMIS:

Example 4 of METS in SIP descriptor:

-<structMap>
*- *
-<fptr FILEID="FILE-0"/>
*- *
-</structMap>

Explanation of Example 4 of METS in SIP descriptor:

"We have a METS structMap. The structMap has one f pointer which references file 0."
--Developer A During Interview with Developer B

Explanation of how the FDA would represent in PREMIS the information in Example 4 of METS in SIP descriptor:

"The original object of type representation is very similar to the structMap or analogous to it. So what we did was we added a relationship in that original representation object to reference file 0, and we tied this back into the structMap. The structMap has one f pointer which references file 0."
--Developer A During Interview with Developer B

The preceding exchanges between the interviewer and developer illustrate the type of adaptation that occurs during the PREMIS implementation process at the FDA. Field observations reveal that after decisions about which entities and semantic units to use have been made, the PREMIS implementation process then involves making prototypes, sharing prototypes with colleagues, testing whether or not those prototypes capture information of interest, and beginning again, if in fact the prototype does not capture the information needed. Field observations suggest that the PREMIS implementation process involves adaptation, and this adaptation involves many steps that are deeply iterative in nature.

For example, during a FDA meeting, Developer A and Developer B shared a prototype of the validation service, which included a series of PREMIS entities and semantic units, the culmination of which completes either successful validation or rejection of SIPs into the FDA. While the prototype was being tested and explained to the entire FDA working group, the FDA director pointed out that the PREMIS entity "Agent" should link to agent description information, not the agent itself. This was different from what had been designed for the prototype example, in which the agent that was being linked to the virus check event was the validation service, when it should have been linked to description information about the virus checker itself.

In this example, there is clear evidence that a prototype had been created, because the prototype was tested during observation. The interviewer could see, from the prototype demonstration, that a number of PREMIS entities and semantic units were being used in the prototype. Discussion during observation highlighted that there were discrepancies between the information that was captured and what the FDA wanted to capture within specific semantic units. This problem did not require new information gathering about semantic units to use, but did require developers to modify the prototype in order to capture the appropriate information for the appropriate semantic units.

In summary, data suggest an adaptation stage as part of the PREMIS implementation process that can be defined or typified by steps. Interview data suggest that using the PREMIS Data Model and Data Dictionary required adaptation at the FDA. Specifically, interview data provide information about how the FDA has had to adapt its own models and concepts to PREMIS. When asked informally about whether there were any challenges, issues, and compromises that had to be dealt with in order to use PREMIS, one of the FDA's lead developers brought up issues concerning decisions that had to be made in order to represent in PREMIS the information they collect about digital objects. This suggests that, from the onset, developers at the FDA have had to make adjustments in order to use PREMIS. Developers at the FDA had to make decisions about how to represent information using PREMIS XML schema that the FDA currently represents using other metadata schemes. Interview data also reveal some clearly defined steps involved in selection of PREMIS entities and semantic units during the PREMIS implementation process. Observational data support additional steps involved in the PREMIS implementation process and also suggest that the nature of the PREMIS implementation process is iterative.

V. Discussion

Analysis of interview and observational data suggest that the PREMIS implementation process at the FDA is largely consistent with DOI and MIS models for implementation and adaptation. Developers' responses during interviews are consistent with Rogers' redefining/restructuring stage in which an innovation, in this case PREMIS, is "re-invented so as to accommodate the organization's needs and structure more closely" (Rogers, 2003, p. 424), while, at the same time, the organization, in this case the FDA, also changes as a result of adjusting in order to use the innovation (Van de Ven, 1986; Leonard-Barton, 1988). Two developers explained 1) that there is a process in which information about PREMIS is compared with the information they have and need to know about their digital objects, and 2) that this process guides how PREMIS is used at the FDA. The fact that the FDA is in the process of creating web services for an updated version of DAITSS 2.0 while incorporating PREMIS 2.0 indicates that the FDA has to make changes in order to use PREMIS (Fischer et al., 2008). Supposing that changes to DAITSS did not stem from difficultly in using PREMIS, the findings support the idea that, from the onset, the FDA had to adapt to use PREMIS. For example, when Developer A stated,

> *"PREMIS has no concept of a package or an intellectual entity, just representations, so the best thing we can do to identify the package in terms of PREMIS is to make a PREMIS object of type representation."*

the developer acknowledged that PREMIS was not a perfect fit for the FDA. Information pertaining to the SIPs received by the FDA had to be reconciled in terms of how they needed to be represented using PREMIS. The interview data suggest that there is a period of information gathering, in which developers compare the FDA's systems, practices, and data with the PREMIS Data Model and Data Dictionary, and then decide how to use the PREMIS Data Model and Data Dictionary to represent the FDA's holdings in its digital repository management system. Thus, the PREMIS implementation process involves adaptation of PREMIS to the FDA and vice versa.

Interview data are consistent with Rogers' (2003) agenda-setting and matching stages, Cooper & Zmud's (1990) initiation stage, and Leonard-Barton's (1988) idea generation and problem solving stages. Rogers' agenda-setting stage involves problem identification, and his matching stage involves fitting an innovation to a predefined problem (Rogers, 2003). Cooper & Zmud's initiation stage is a process in which "active and/or passive scanning of organizational problems/opportunities and IT solutions are undertaken" (1990, p. 124). Leonard-Barton's (1988) idea generation and problem solving stages involve trying to find a solution to a problem, based on specifics regarding an organization and the innovative technology considered, for use within a given organization (pp. 260-261). All of these stages are comparable to each other in that they involve members of organizations, responsible for implementing an innovation, comparing the organization to the innovation to address a particular problem or set of issues.

When asked if the PREMIS implementation process could be defined or typified by steps, developers responded by defining steps that are strikingly similar to those previously identified by Rogers, Cooper & Zmud, and Leonard-Barton. For example, Developer A stated:

> *[w]e start with our problem, from the problem, we get the information we know we need to represent, and then once we have that information, we choose the appropriate semantic units of PREMIS.*
>
> *--Developer A*

In this example, the developer describes a process in which developers gather information about the FDA (its systems, practices, and data) and information about PREMIS (its Data Model and Data Dictionary), compare all of this information, and then make judgment calls – the outcomes of which involve selection of appropriate PREMIS semantic units and entities, based on the FDA's information need for proper representation of information needed for long term preservation and access to its holdings. Based on this interview data, the PREMIS implementation process first involves information gathering.

Observational data are consistent with Rogers' (2003) redefining/restructuring stage, Cooper & Zmud's (1990) adaptation stage, and Leonard-Barton's (1988) small cycle of redefinition when developing technology. Rogers' redefining/restructuring stage is a process in which an organization redefines and restructures both the technology and

organizational processes (Rogers, 2003). Cooper & Zmud's adaptation stage is a process in which the IT application is developed, installed, and maintained; organizational procedures are revised and developed; and organizational members are trained both in the new procedures and in the IT application (Cooper & Zmud, 1990, p. 124). Leonard-Barton's (1988) small cycle of redefinition when developing technology involves creation of a laboratory prototype which leads to a pilot production prototype, which leads to a production prototype, which leads to production-ready technology. If a group of developers finish a laboratory prototype, test the laboratory prototype in the pilot production prototype stage, and find that the prototype is somehow defective, the group must reconvene, go back to the laboratory prototype stage, try to resolve problem(s), and then once more enter the pilot production prototype stage. If errors occur during the production prototype stage, the developers likewise have to go back to the laboratory prototype stage and try to resolve the problematic issues. Then developers must head back to the pilot production prototype stage – with success – before again returning to the production prototype stage.

At the FDA, adaptation is executed similarly to Leonard-Barton's small cycle of redefinition when developing technology. Observation of the prototype demonstration at the FDA revealed that a large number of PREMIS entities and semantic units were being juggled simultaneously and iteratively during implementation. Discussion during observation highlighted that there were discrepancies between the information that was captured by the prototype and what the FDA wanted to capture within specific semantic units. This problem did not require new information gathering about semantic units to use, but instead required developers to modify the prototype in order to capture the appropriate information for the appropriate semantic units. This modification process is inherent in Leonard-Barton's model and is made explicit by arrows pointing back to previous stages if prototypes are found deficient in any particular adaptation stage.

Based on the data gathered during the study, a modification of the standard and well-documented DOI/MIS models is proposed. The Iterative Model for the Adaptation Stage of the PREMIS Implementation Process (See Figure 1) includes four distinctive stages: Information Gathering; Prototyping; Sharing, and Revision or Implement. The *Information Gathering Stage* is comparable to Rogers' (2003) agenda-setting and matching stages, Cooper & Zmud's (1990) initiation stage, and Leonard Barton's (1988) idea generation and problem solving stages in which members of organizations, responsible for implementing an innovation, compare their organization to a particular innovation in order to address a particular problem or set of issues. The *Prototyping Stage* is directly derived from Leonard-Barton's (1988) laboratory prototype stage in which prototypes are created and the *Sharing Stage* is directly related to Leonard-Barton's (1988) pilot production prototype stage in which the prototype

Figure 1. Iterative Model for the Adaptation Stage of the PREMIS Implementation Process Adapted from Rogers (2003) and Leonard-Barton (1988)

Insert Figure 1

is tested on a larger scale and shared among colleagues or otherwise distributed through a system or organization for large-scale testing. The *Revision or Implement Stage* most strongly relates to Leonard-Barton's (1988) model, in which the arrows pointing from production prototype to previous stages, i.e. laboratory prototype and then pilot production prototype, denote the need to revise, and the progression toward the last stage, production-ready technology, denotes the qualification to move forward with a prototype to full implementation – if, and only if, there are no problems with the prototype.

In the example highlighted earlier, in which the developers conducted prototype testing during a FDA meeting, and the FDA Director pointed out that the PREMIS entity *Agent* should link to agent description information, not the agent itself, the Information Gathering Stage and Prototyping Stage had already been completed before this study began, the researchers witnessed the Sharing Stage as well as the Revision or Implement Stage. The outcome of the Sharing Stage involved moving to the Revision or Implement Stage in which the decision to revise was deemed the appropriate action to take.

Thus, data collected during this study suggest that the FDA experienced each step of the Iterative Model for the Adaptation Stage of the PREMIS Implementation Process in the following ways; during the Information Gathering Stage, the FDA gathered information from the PREMIS Data Dictionary (definitions, rationale, examples) about entities and semantic units, and decided which of them, if used, could fulfill preservation metadata information needs. After gathering enough information, the FDA moved to the Prototyping Stage, in which developers created prototypes for gathering the information required for each chosen PREMIS semantic unit and entity. After creating a prototype, developers at the FDA shared prototypes with colleagues and senior management and got feedback during the Sharing Stage. Next, during the Revision or Implement Stage, if revision was necessary, developers at the FDA repeated the preceding steps – with the exception of the Information Gathering Stage. If revision was not necessary, senior management at the FDA allowed implementation decisions inherent in the prototype to continue through the implementation process.

VI. Conclusion

This case study explores what the adaptation stage looks like when implementing PREMIS with careful attention toward how developers describe this process. This exploration used the Florida Digital Archive as an exemplar of system implementation, because the FDA is deeply engaged with the PREMIS implementation process and has at its disposal a variety of resources that support successful implementation. The study constructed an in-depth picture of the PREMIS implementation process at FDA by comparing developers' answers to questions with observed behavior as well as comparing all data collected with innovation adoption models suggested by the DOI and MIS literature. Analysis of data collected during interviews and observations suggest that, at the Florida Digital Archive, an adaptation stage clearly exists as part of the PREMIS implementation process, one that can be defined or typified by successive, iterative steps.

Specifically, the case study of the FDA suggests:

1) that use of the PREMIS Data Model and Data Dictionary requires adaptation;

2) clearly identifiable steps exist in the selection of PREMIS entities and semantic units during the PREMIS implementation process;

3) there are additional steps involved in the PREMIS implementation process after choices about entities and semantic units have been made; and

4) the nature of the PREMIS implementation process is iterative.

Based on DOI/MIS literature and models for implementation and data collected at the FDA while developers were in the process of implementing PREMIS, this study proposes an Iterative Model for the Adaptation Stage of the PREMIS Implementation Process. Although existing general models for innovation adaptation can be useful for understanding the PREMIS implementation process, the proposed model specifically addresses the implementation process in the context of PREMIS.

The extent to which the Iterative Model for the Adaptation Stage of the PREMIS Implementation Process is generalizable beyond the Florida Digital Archive setting remains to be tested. Future studies of PREMIS implementation could benefit from more varied interviews and longer observational periods. One of the main goals of this study was to get digital preservationists thinking about the PREMIS implementation process, because of its importance. The digital preservation community could benefit from future studies which could test the validity of the Iterative Model for the Adaptation Stage of the PREMIS Implementation Process using the FDA again as well as other institutions currently implementing PREMIS.

This study is an exercise in understanding the nature of preservation work from the perspective of systems programmers who themselves may or may not have any abiding commitment to digital preservation. On the surface, it may seem as though developers who implement a preservation metadata framework are only engaged in making a series of innocuous decisions that do not really have anything to do with the grander and perhaps more abstract goals of digital preservation. A closer look at the PREMIS implementation process, in particular, the adaptation stage of the PREMIS implementation process, reveals that individual decisions made by systems developers have major implications for long-term preservation of digital content. Choices about how to reconcile a digital repository's systems, practices, and models with the PREMIS Data

Model and Data Dictionary determine what information is considered, kept, maintained, and linked to the digital objects the FDA is responsible for safekeeping. Because digital objects will not preserve themselves and digital preservation management systems will not adapt themselves to models and data dictionaries for preservation, successful digital preservation devolves to the highly detailed decision making that is required to implement preservation standards. And so it is that preservation is constructed, rather than willed into existence – such construction being a detailed, iterative process of molding and tweaking existing systems, practices, models, processes and information to synch with temporally constrained standards, data models, and data dictionaries. In this way, preservation standards support preservation from a conceptual perspective, but in and of themselves, are incapable of affecting preservation.

References

Alemneh, D. (2008), "Barriers to Adopting PREMIS in Cultural Heritage Institutions: an Exploratory Study", in *Archiving 2008 Final Program and Proceedings*, Washington, D.C., pp. 71-80.

Berman, P. (1980), "Thinking About Programmed and Adaptive Implementation: Making Strategies to Situations", in Ingram, H. and Mann, D. (Eds.), *Why Policies Succeed or Fail,* Sage, Beverly Hills, CA, pp. 200-220.

Caplan, P. and Guenther, R. (2005), "Practical Preservation: The PREMIS Experience", *Library Trends*, Vol. 54 No. 1, pp. 111-124.

Caplan, P. (2009), "Understanding PREMIS", available at: http://www.loc.gov/standards/premis/understanding-premis.pdf (accessed 12 September 2009).

CCSDS (2002), "Reference model for an Open Archival Information System (OAIS)", Consultative Committee on Space Data Systems, Washington, D.C.

Conway, P. (1996), "Preservation in the Digital World", Commission on Preservation and Access, Washington, DC, available at: http://www.clir.org/pubs/reports/conway2/ (accessed 11 October 2009).

Cooper, R. B., and Zmud, R.W. (1990), "Information Technology Implementation Research: A Technological Diffusion Approach", *Management Science*, Vol. 36, pp. 123-139.

Creswell, J. W. (2007), *Qualitative Inquiry & Research Design: Choosing Among Five Approaches* second edition, Sage Publications, Thousand Oaks, CA.

Dappert, A. and Enders, M. (2008), "Using METS, PREMIS and MODS for Archiving eJournals", *D-Lib Magazine,* available at: http://www.dlib.org/dlib/september08/dappert/09dappert.html (accessed 12 October 2009).

Fischer, R., Chou, C., and Lazzarino, F. (2008), "Updating DAITSS – Transitioning to a web service architecture", paper presented at iPres, 29-30 September, London, UK, available at: http://www.bl.uk/ipres2008/presentations_day2/35_Fischer.pdf (accessed 22 October 2009).

Florida Center for Library Automation (2006), "DAITSS and PREMIS Conformance", available at: http://daitss.fcla.edu/wiki/DocumentationPage (accessed 22 October 2009).

Florida Digital Archive (2009), "Monthly Report of Total SIPs/Files/Bytes Ingested between Sep 1, 2009 12:00 AM and Oct 1, 2009 12:00 AM", available at: http://fclaweb.fcla.edu/uploads/Lydia%20Motyka/FDA_Statistics/monthly_ingest_Sept_09.pdf (accessed 1 November 2009).

----- (2009), "A proposed change in Preservation Level in FDA Agreements, Sept. 22, 2008", available at: http://blogs.fcla.edu/index.php/digitalarchive?s=normalization&submit=Search (accessed 1 November 2009).

----- (2009), "DAITSS 2.0 Service Architecture: Ingest", available at: http://daitss.fcla.edu/wiki/DaitssTwo (accessed 1 November 2009).

----- (2006), "Submitting Materials to the FCLA Digital Archive", available at: http://www.fcla.edu/digitalArchive/pdfs/SubmittingMaterials.pdf (accessed 1 November 2009).

Garrett, J., and Waters, D. (1996), "Preserving digital information", available at: http://www.oclc.org/research/activities/past/rlg/digpresstudy/final-report.pdf (accessed 17 December 2009).

Guenther, R. and Xie, Z., (2007), "Implementing PREMIS in Container Formats", paper presented at Archiving 2007, 21- 24 May, Arlington, Virginia, available at: http://www.loc.gov/standards/premis/IST-premis-containers.pdf (accessed 14 December 2009).

Guenther, R. (2008), "Guidelines for using PREMIS with METS for exchange", available at: http://www.loc.gov/standards/premis/guidelines-premismets.pdf (accessed 5 December 2009).

Hedstrom, M. (2003), "It's About Time: Research Challenges in Digital Archiving and Long-term Preservation: Final Report on the NSF Workshop on Research Challenges in Digital Archiving and Long-term Preservation", 12-13April, 2002, Washington, D.C.

Knight, G., & Hedges, M. (2007), "Modeling OAIS Compliance for Disaggregated Preservation Services", *International Journal of Digital Curation*, Vol. 1 No. 2, pp.62-72.

Lavoie, B. (2004), "Implementing Metadata in Digital Preservation Systems: The PREMIS Activity", *D-Lib Magazine*, available at: http://www.dlib.org/dlib/april04/lavoie/04lavoie.html (accessed 12 December 2009).

Lavoie, B. and Gartner, R. (2005), "Technology Watch Report: Preservation Metadata", available at: http://www.dpconline.org/docs/reports/dpctw05-01.pdf (accessed 1 October 2009).

Lee, B., Clifton, G., and Langley, S. (2006), "PREMIS Requirement Statement Project Report. Australian Partnership for Sustainable Repositories", available at: http://www.apsr.edu.au/publications/presta.pdf (accessed 1 January 2009).

Leonard-Barton, D. (1988), "Implementation as mutual adaptation of technology and Organization", *Research Policy*, Vol. 17 No. 5, pp. 251-267.

McCallum, S. H. (2005), "Preservation Metadata Standards for Digital Resources: what we have and what we need", paper presented at *World Library and Information Congress: 71th IFLA General Conference and Council*, available at: http://www.ifla.org/IV/Ifla71/papers/060e-McCallum.pdf (accessed 23 September 2009).

OCLC/RLG Working Group on Preservation Metadata (2002), "Preservation Metadata and the OAIS Information Model: A Metadata Framework to Support the Preservation of Digital Objects", available at: http://www.oclc.org/research/activities/past/orprojects/pmwg/pm_framework.pdf (accessed 1 December 2009).

----- (2001), "Preservation Metadata for Digital Objects: A Review of the State of the Art", available at: http://www.oclc.org/research/activities/past/orprojects/pmwg/presmeta_wp.pdf (accessed 22 October 2009).

PADI Working Group (2007), "Preservation Metadata." available at: http://www.nla.gov.au/padi/topics/32.html (accessed 13 January 2009).

Pearce, J., Pearson, D., Williams, M., and Yeadon, S. (2008), "The Australian METS Profile – A Journey about Metadata", *D-Lib Magazine*, available at: http://www.dlib.org/dlib/march08/pearce/03pearce.html (accessed 11 October 2009).

PREMIS Editorial Committee (2008), "PREMIS data dictionary for Preservation Metadata."
available at: http://www.loc.gov/standards/premis/v2/premis-2-0.pdf (accessed 13 January 2009).

PREMIS Maintenance Activity (2010), "PREMIS Implementation Registry", available at: http://www.loc.gov/standards/premis/premis-registry.php (accessed 31 January 2010).

PREMIS Working Group (2004), "Implementing Preservation Repositories For Digital Materials: Current Practice and Emerging Trends In The Cultural Heritage Community", available at: http://www.oclc.org/research/projects/pmwg/surveyreport.pdf (accessed 13 January 2009).

-----. (2005), "Data Dictionary for Preservation Metadata", available at: http://www.oclc.org/research/activities/past/orprojects/pmwg/premis-final.pdf (accessed 22 October 2009).

Rogers, E. M. (2003), *Diffusion of innovations,* 5th ed., Free Press, New York, NY.

Seadle, M. (2009), "Archiving in the networked world: betting on the future", *Library Hi Tech,* Vol. 27 No. 2, pp. 319-325.

Society of American Archivists (2006), "Preservation Publication Award", available at: http://www.archivists.org/recognition/dc2006-awards.asp#preservation (accessed 20 August 2009).

Van de Ven, A. H. (1986), "Central problems in the management of innovation", *Management Science*, Vol. 32 No. 5, pp. 590-607.

Woodyard-Robinson, D. (2007), "Implementing the PREMIS data dictionary: a survey of Approaches", available at: http://www.loc.gov/standards/premis/implementation-report-woodyard.pdf (accessed 1 October 2009).

Yin, R. K. (2003), *Case study research: Design and method*, third edition, Thousand Oaks, CA.

122x57mm (150 x 150 DPI)